Assessing Uncertainty in High-dimensional Regression Models

Chen Shizhe

Department of Biostatistics
University of Washington

April 18, 2013

Papers:

- Statistical significance in high-dimensional linear models, Bühlmann (2012).
- On asymptotically optimal confidence regions and tests for high-dimensional models, van de Geer, Bühlmann, and Ritov (2013).

Associations: Marginal v.s. Conditional

Genes: $\{A^*\}$ $\{B_1^*, B_2, B_3\}$ $\{C_1, C_2, C_3\}$

(An example from Grazier G'Sell et al. 2013.)

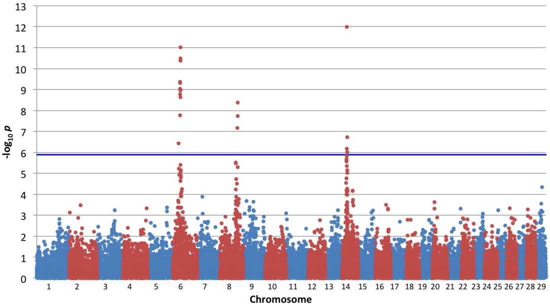

Figure: A Manhattan plot from Nishimura et al. (2012).

Regression models on high-dimensional data

$$Y = X\beta^* + \epsilon, \quad \epsilon \sim N_n(0, I_n). \tag{1}$$

OLS/ML on this simple model is equivalent to:

$$\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \; \|Y - X\beta\|_2^2. \tag{2}$$

Linear algebra gives $\hat{\beta} = (X^T X)^{-1} X^T Y$.

However, $X^T X$ is non-invertible when $p > n$. Consider the ridge regression:

$$\underset{\beta \in \mathbb{R}^p}{\text{minimize}} \; \|Y - X\beta\|_2^2 + \lambda_n \sum_{i=1}^{p} \beta_i^2. \tag{3}$$

Penalized regression

There are many choices of penalties:

- ℓ_2 norm, i.e. $\|\beta\|_2^2 = \sum_{i=1}^{p} \beta_i^2$, ridge regressions.
- ℓ_1 norm, i.e. $\|\beta\|_1 = \sum_{i=1}^{p} |\beta_i|$, lasso regressions.
- and more.

Advantages of using certain penalties:

1. Each penalty imposes a low-dimensional structure.
2. Convex loss functions.

Assessing uncertainty

- Lasso: "Sparse estimators such as Lasso do not have a tractable limiting distribution. The limiting distributions of Lasso estimators depend on unknown parameters in low-dimensional setting. Bootstrap and subsampling techniques are plagued by non-continuity of limiting distributions." (van de Geer et al., 2013).

- Ridge regression: $\text{Bias}(\hat{\beta}) = -\lambda_n(X^TX + \lambda_n I)^{-1}\beta$; Bootstrap in low-dimensional setting, see Crivelli et al. (1995).

Related work

- Variable selections: Correct recover of support requires the "beta-min" assumption (Meinshausen and Bühlmann 2006, Wainwright 2009, Negahban et al. 2010).
- Bootstrap and subsampling method. In low-dimensional setting, bootstraps for lasso are proposed by Chatterjee and Lahiri (2011), Sartori (2011). For Ridge regression, see Crivelli et al. (1995).
- A significance test for the lasso (Lockhart, Taylor, Tibshirani, and Tibshirani, 2013).

Another line of research...

- Zhang and Zhang (2011), and Bühlmann (2012) provided hypothesis testing procedures on high-dimensional linear models.
- Javanmard and Montanari (2013) provided a minimax test for linear models.
- van de Geer et al. (2013) claimed to reach the semiparametric efficiency bound, and the method be generalized to ℓ_1-penalized GLMs and ridge regressions.

Summary

- Marginal associations v.s. conditional associations.
- Reasons for using penalized regressions on high-dimensional data.
- Current attempts to make statistical inference on high-dimensional regressions.

A glimpse ahead

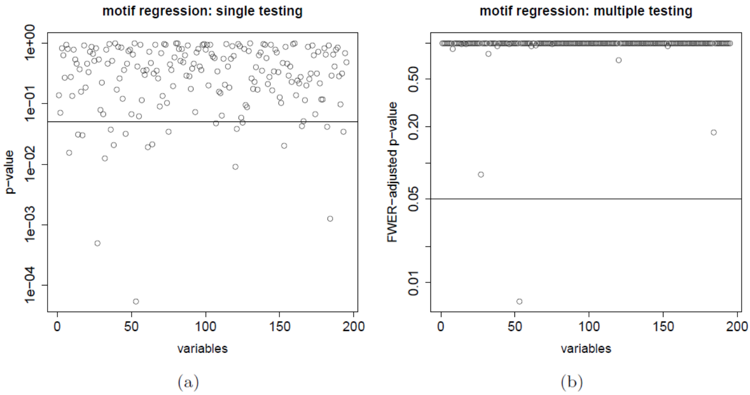

Figure: A Motif regression result from Bühlmann (2012).

The method in van de Geer et al. (2013)

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}}(\|Y - X\beta\|_2^2/(2n) + \lambda\|\beta\|_1). \tag{4}$$

The Karush-Kuhn-Tucker conditions are

$$\begin{aligned} &-X^T(Y - X\hat{\beta}) + \lambda\hat{\tau} = 0, \\ &\|\hat{\tau}\|_\infty \leq 1, \text{ and } \hat{\tau}_j = \text{ sign } (\hat{\beta}_j) \text{ if } \hat{\beta}_j \neq 0. \end{aligned} \tag{5}$$

Using the first equation, we have

$$n^{-1}X^TX(\hat{\beta} - \beta^*) + \lambda\hat{\tau} = X^T\epsilon/n. \tag{6}$$

Now assume we have a $\hat{\Theta}$ that is a "relaxed form" of an inverse of $n^{-1}X^TX$.

$$\hat{\beta} - \beta^* + \hat{\Theta}\lambda\hat{\tau} = \hat{\Theta}X^T\epsilon/n - \Delta, \tag{7}$$

where $\Delta = (\hat{\Theta}\hat{\Sigma} - I)(\hat{\beta} - \beta)$.

The method in van de Geer et al. (2013), continued

We will show that Δ, in fact $\sqrt{n}\Delta$, is asymptotically negligible under certain assumptions on sparsity. Then let

$$\hat{b} = \hat{\beta} + \hat{\Theta} X^T (Y - X\hat{\beta})/n. \tag{8}$$

Theorem 2.2 in van de Geer et al. (2013) claims that:

$$\sqrt{n}(\hat{b} - \beta^*) = W + o_P(1), \quad W|X \sim N_p(0, \sigma_\epsilon^2 \hat{\Theta}\hat{\Sigma}\hat{\Theta}^T). \tag{9}$$

Reference I

Peter Bühlmann. Statistical significance in high-dimensional linear models. *arXiv preprint arXiv:1202.1377*, 2012.

Arindam Chatterjee and Soumendra Nath Lahiri. Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494), 2011.

Ana Crivelli, Luis Firinguetti, Rosa Montaño, and Margarita Muñóz. Confidence intervals in ridge regression by bootstrapping the dependent variable: a simulation study. *Communications in Statistics-Simulation and Computation*, 24(3):631–652, 1995.

Max Grazier G'Sell, Trevor Hastie, and Robert Tibshirani. False variable selection rates in regression. 2013.

Adel Javanmard and Andrea Montanari. Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory. *arXiv preprint arXiv:1301.4240*, 2013.

Richard Lockhart, Jonathan Taylor, Ryan Tibshirani, and Robert Tibshirani. A significance test for the lasso. *arXiv preprint arXiv:1301.7161*, 2013.

N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.

Sahand Negahban, Pradeep Ravikumar, Martin J Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of *m*-estimators with decomposable regularizers. *arXiv preprint arXiv:1010.2731*, 2010.

Shota Nishimura, Toshio Watanabe, Kazunori Mizoshita, Ken Tatsuda, Tatsuo Fujita, Naoto Watanabe, Yoshikazu Sugimoto, and Akiko Takasuga. Genome-wide association study identified three major qtl for carcass weight including the plag1-chchd7 qtn for stature in japanese black cattle. *BMC genetics*, 13(1):40, 2012.

Samantha Sartori. *PENALIZED REGRESSION: BOOTSTRAP CONFIDENCE INTERVALS AND VARIABLE SELECTION FOR HIGH-DIMENSIONAL DATA SETS*. PhD thesis, Università degli Studi di Milano, 2011.

Sara van de Geer, Peter Bühlmann, and Ya'acov Ritov. On asymptotically optimal confidence regions and tests for high-dimensional models. *arXiv preprint arXiv:1303.0518*, 2013.

M.J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using¡ formula formulatype=. *Information Theory, IEEE Transactions on*, 55(5):2183–2202, 2009.

Cun-Hui Zhang and S Zhang. Confidence intervals for low-dimensional parameters with high-dimensional data. *arXiv preprint arXiv:1110.2563*, 2011.